Triplet Mining-Based Phishing Webpage Detection

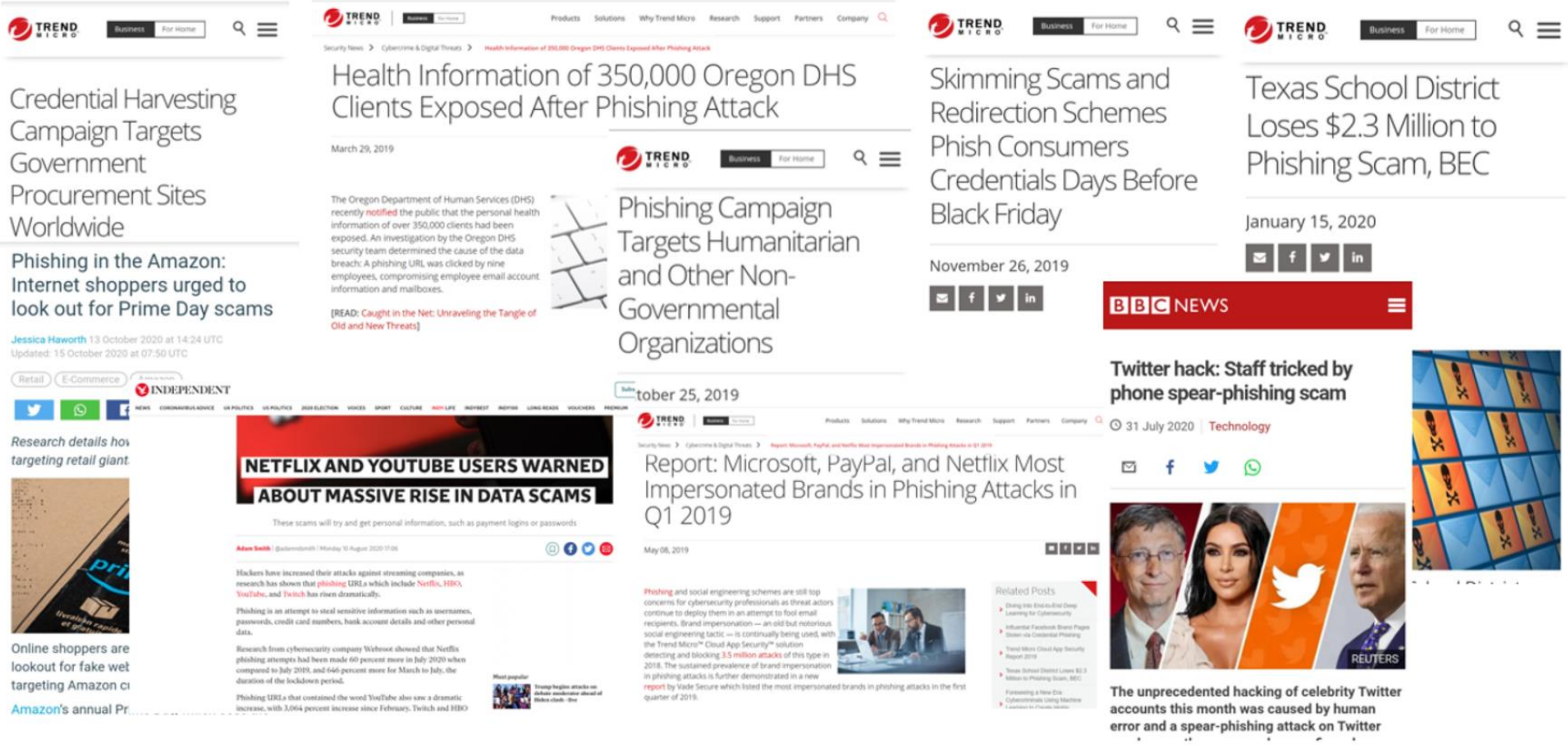
Kalana Abeywardena, Jiawei Zhao, Lexi Brent, Suranga Seneviratne, and Ralph Holz

THE UNIVERSITY OF SYDNEY

विद्यैव संवर्धनम् ।
UNIVERSITY OF MORATUWA

University of Twente
The Netherlands

Motivation



Credential Harvesting Campaign Targets Government Procurement Sites Worldwide

Phishing in the Amazon: Internet shoppers urged to look out for Prime Day scams

Jessica Haworth 13 October 2020 at 14:24 UTC
Updated: 15 October 2020 at 07:50 UTC

Retail E-Commerce

Research details hov targeting retail giant

Online shoppers are lookout for fake wel targeting Amazon cı

Amazon's annual Pr

Security News > Cybercrime & Digital Threats > Health Information of 350,000 Oregon DHS Clients Exposed After Phishing Attack

Health Information of 350,000 Oregon DHS Clients Exposed After Phishing Attack

March 29, 2019

The Oregon Department of Human Services (DHS) recently notified the public that the personal health information of over 350,000 clients had been exposed. An investigation by the Oregon DHS security team determined the cause of the data breach: A phishing URL was clicked by nine employees, compromising employee email account information and mailboxes.

[READ: Caught in the Net: Unraveling the Tangle of Old and New Threats]

Phishing Campaign Targets Humanitarian and Other Non-Governmental Organizations

tober 25, 2019

Skimming Scams and Redirection Schemes Phish Consumers Credentials Days Before Black Friday

November 26, 2019

Texas School District Loses $2.3 Million to Phishing Scam, BEC

January 15, 2020

INDEPENDENT

NETFLIX AND YOUTUBE USERS WARNED ABOUT MASSIVE RISE IN DATA SCAMS

These scams will try and get personal information, such as payment logins or passwords

Adam Smith | @adamndsmith | Monday 10 August 2020 17:06

Hackers have increased their attacks against streaming companies, as research has shown that phishing URLs which include Netflix, HBO, YouTube, and Twitch has risen dramatically.

Phishing is an attempt to steal sensitive information such as usernames, passwords, credit card numbers, bank account details and other personal data.

Research from cybersecurity company Webroot showed that Netflix phishing attempts had been made 60 percent more in July 2020 when compared to July 2019, and 646 percent more for March to July, the duration of the lockdown period.

Phishing URLs that contained the word YouTube also saw a dramatic increase, with 3,064 percent increase since February. Twitch and HBO

Most popular

Trump begins attacks on debate moderator ahead of Biden clash - live

Security News > Cybercrime & Digital Threats > Report: Microsoft, PayPal, and Netflix Most Impersonated Brands in Phishing Attacks in Q1 2019

Report: Microsoft, PayPal, and Netflix Most Impersonated Brands in Phishing Attacks in Q1 2019

May 08, 2019

Phishing and social engineering schemes are still top concerns for cybersecurity professionals as threat actors continue to deploy them in an attempt to fool email recipients. Brand impersonation — an old but notorious social engineering tactic — is continually being used, with the Trend Micro™ Cloud App Security™ solution detecting and blocking 3.5 million attacks of this type in 2018. The sustained prevalence of brand impersonation in phishing attacks is further demonstrated in a new report by Vade Secure which listed the most impersonated brands in phishing attacks in the first quarter of 2019.

Related Posts

- Diving Into End-to-End Deep Learning for Cybersecurity
- Influential Facebook Brand Pages Stolen via Credential Phishing
- Trend Micro Cloud App Security Report 2019
- Texas School District Loses $2.3 Million to Phishing Scam, BEC
- Forecasting a New Era: Cybercriminals Using Machine Learning to Create Highly

BBC NEWS

Twitter hack: Staff tricked by phone spear-phishing scam

31 July 2020 | Technology

The unprecedented hacking of celebrity Twitter accounts this month was caused by human error and a spear-phishing attack on Twitter

Motivation

YouTube
twitter
unicef
DHS
Oregon Department
of Human Services
PayPal
Microsoft

Phishing Webpage Detection

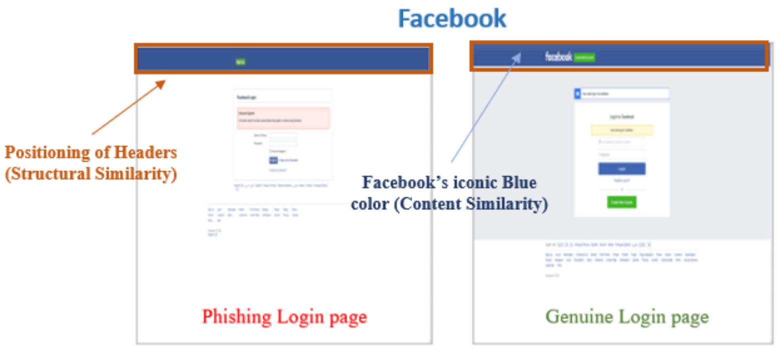
Different Methods currently in use:

- Blacklisting or Whitelisting
- Texual-feature based ML techniques
- Visual Similarity between webpages

Phishing Webpage Detection

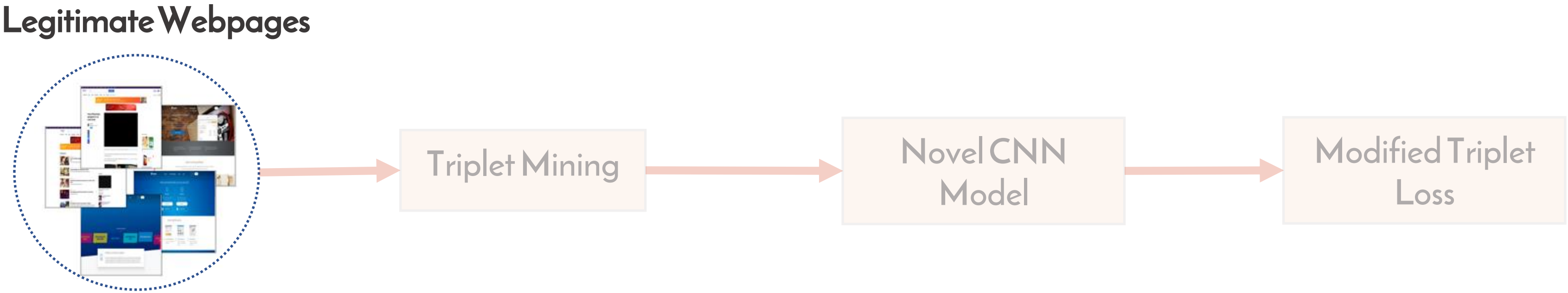
Different Methods currently in use:

- Blacklisting or Whitelisting
- Texual-feature based ML techniques
- **Visual Similarity between webpages**



- Novel CNN Architecture
- Content Similarity – Content Embedding
- Structural Similarity – Structural Embedding
- Triplet Learning WITHOUT Phishing webpages

Training Pipeline



Legitimate Webpages

Triplet Mining

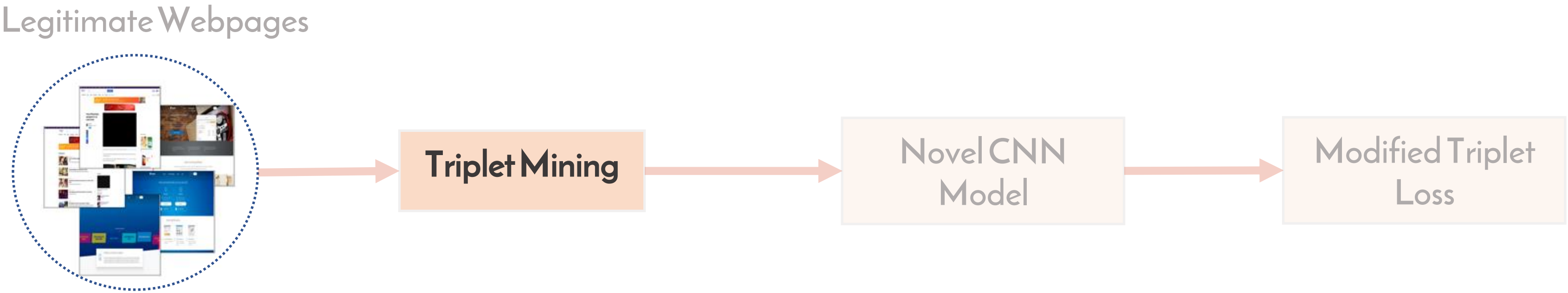
Novel CNN Model

Modified Triplet Loss

01 Data Collection and Preparation

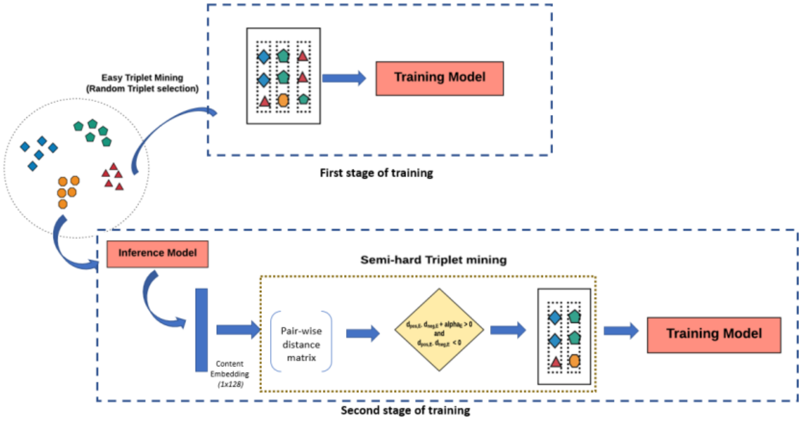
- Scraping Headless chrome pages and saving (1920, 1080, 3) screenshots
- 49063 webpages of 9557 domains including 3619 logging pages
- Split to Train and Validation

Training Pipeline

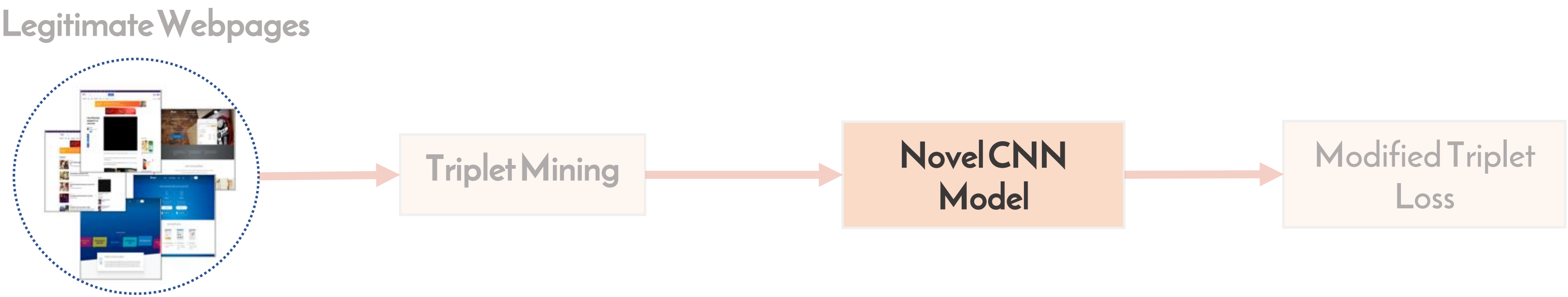


02 | Triplet Mining

- Two available Triplet Mining Strategies used.
- Phase 01 – Easy/Random Triplet Mining
- Phase 02 – Semi-Hard Triplet Mining


Easy Triplet Mining
(Random Triplet selection)
Training Model
First stage of training
Inference Model
Semi-hard Triplet mining
Content Embedding (1x128)
Pair-wise distance matrix
$d_{pos,E} - d_{neg,E} + alpha_E > 0$
and
$d_{pos,E} - d_{neg,E} < 0$
Training Model
Second stage of training

Training Pipeline



Legitimate Webpages

Triplet Mining

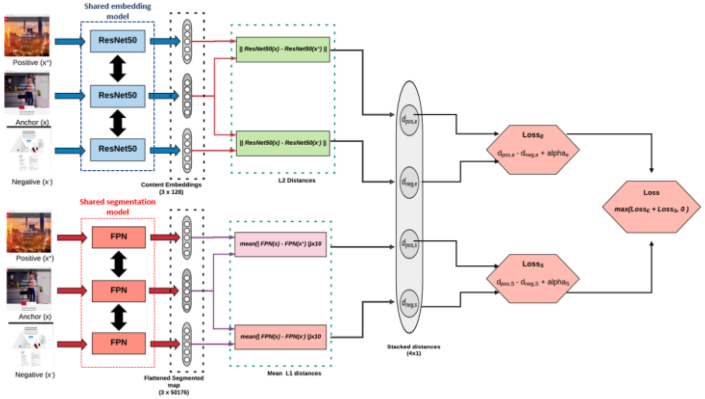
Novel CNN Model

Modified Triplet Loss

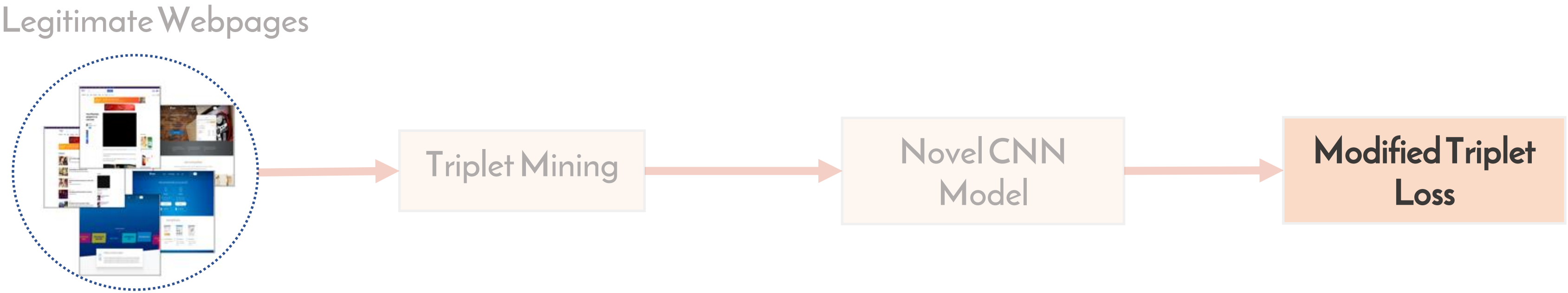
03

Novel CNN Model

- Fusion between two Triplet Networks.
- Triplet Network for Content Similarity Detection
- Triplet Network for Structural Similarity Detection


Shared embedding model
Positive (x^+)
Anchor (x)
Negative (x^-)
ResNet50
ResNet50
ResNet50
Content Embeddings (3 x 128)
|| ResNet50(x) - ResNet50(x⁺) ||
|| ResNet50(x) - ResNet50(x⁻) ||
L2 Distances
Shared segmentation model
Positive (x^+)
Anchor (x)
Negative (x^-)
FPN
FPN
FPN
Flattened Segmented map (3 x 50176)
mean(| FPN(x) - FPN(x⁺) |)x10
mean(| FPN(x) - FPN(x⁻) |)x10
Mean L1 distances
$d_{pos,e}$
$d_{neg,e}$
$d_{pos,s}$
$d_{neg,s}$
Stacked distances (4x1)
$Loss_E$
$d_{pos,e} - d_{neg,e} + alpha_e$
$Loss_S$
$d_{pos,S} - d_{neg,S} + alpha_S$
Loss
$max(Loss_E + Loss_S, 0)$

Training Pipeline



Legitimate Webpages

Triplet Mining

Novel CNN Model

Modified Triplet Loss

04 Modified Triplet Loss

- Uses a modified Triplet Loss based on Triplet losses from two Triplet Networks fused to optimize the weights.

$$loss = \sum_i max(loss_c^{(i)} + loss_s^{(i)}, 0)$$

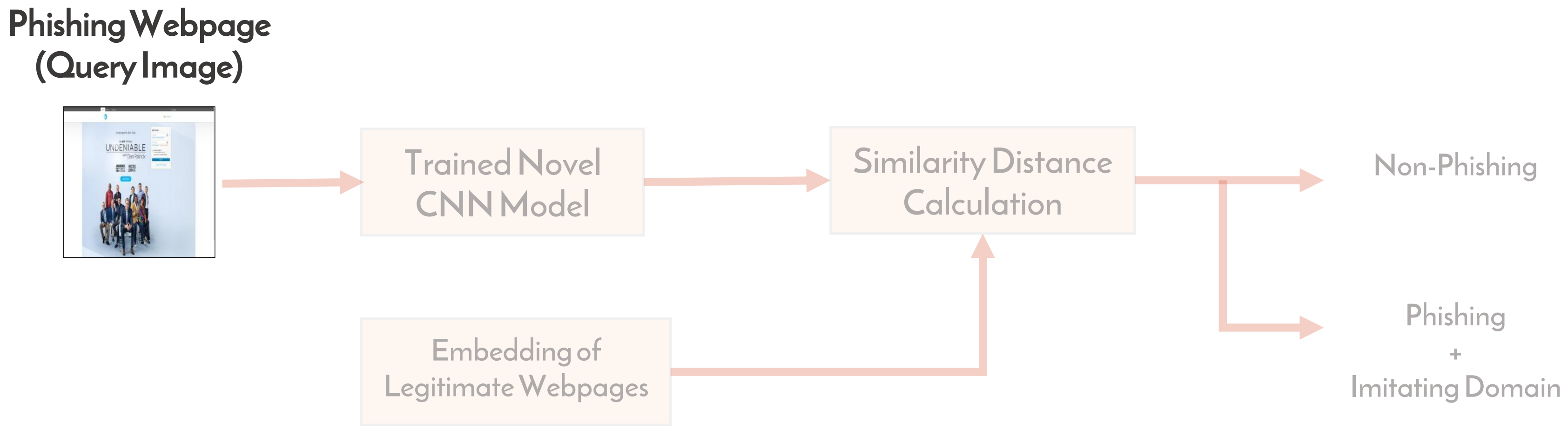
where

$$loss_c^{(i)} = d_{pos,c}^{(i)} + \alpha_c - d_{neg,c}^{(i)}$$

and

$$loss_s^{(i)} = d_{pos,s}^{(i)} + \alpha_s - d_{neg,s}^{(i)}$$

Phishing Webpage Detection Pipeline



Phishing Webpage (Query Image)

Trained Novel CNN Model

Similarity Distance Calculation

Embedding of Legitimate Webpages

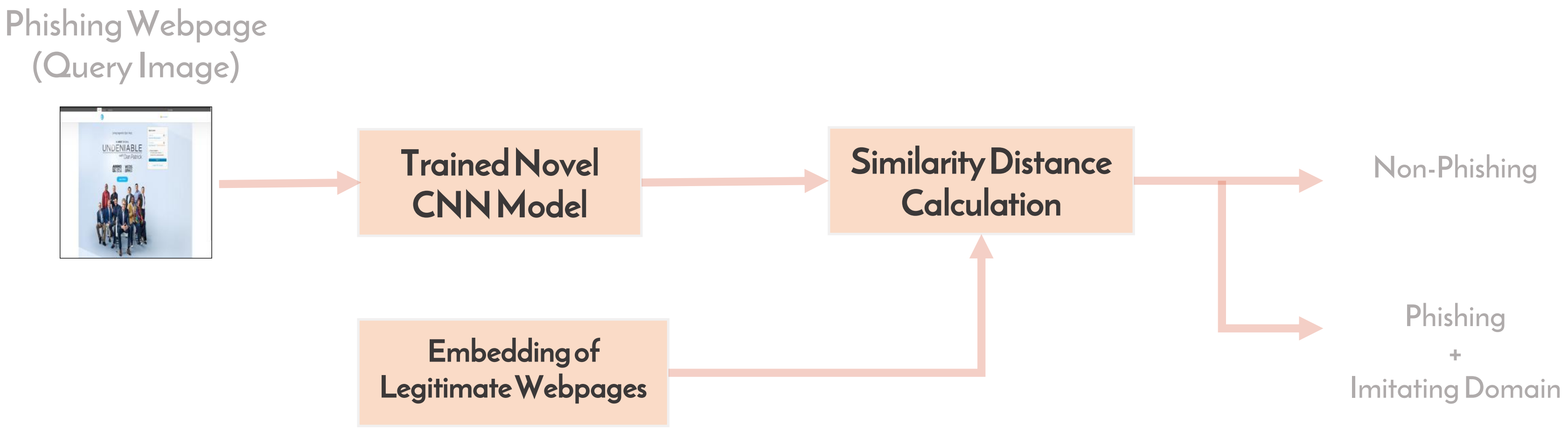
Non-Phishing

Phishing + Imitating Domain

01 Data Collection and Preparation

- Scraping Headless chrome pages of Phishing Webpages on Phishtank (2019-02-01)
- 113 phishing webpages imitating 22 legitimate sites

Phishing Webpage Detection Pipeline



Phishing Webpage
(Query Image)

Trained Novel CNN Model

Similarity Distance Calculation

Embedding of Legitimate Webpages

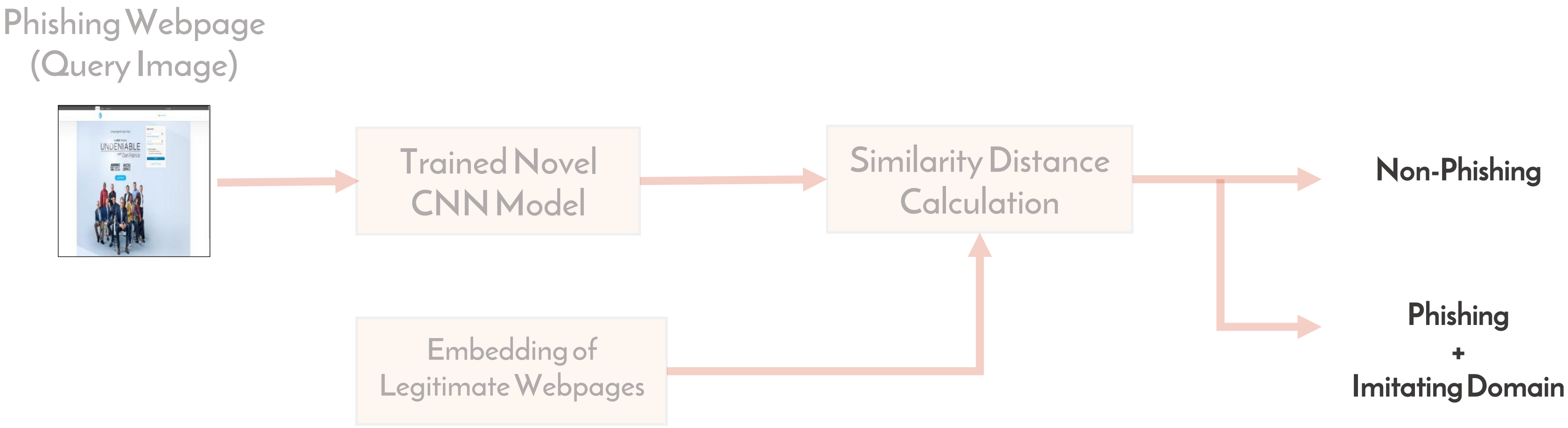
Non-Phishing

Phishing
+
Imitating Domain

02 | Similarity Calculation

- Distance calculation between embeddings of query image (from Trained CNN model) and pre-computed embeddings of legitimate webpages

Phishing Webpage Detection Pipeline



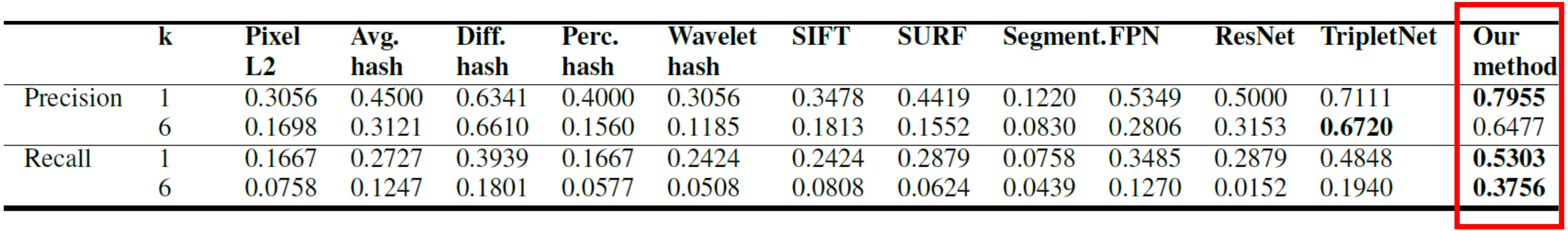
03 Detection

- Based on a pre-calculated threshold value, detects a Query Image as Phishing or Non-Phishing
- If Phishing is detected, returns Top-1 legitimate domain

Experiments and Results

Different Baseline Image Matching methods compared:

- Raw pixel-wise distance
- Hashing methods
- Image embeddings related methods



	k	Pixel L2	Avg. hash	Diff. hash	Perc. hash	Wavelet hash	SIFT	SURF	Segment.	FPN	ResNet	TripletNet	Our method
Precision	1	0.3056	0.4500	0.6341	0.4000	0.3056	0.3478	0.4419	0.1220	0.5349	0.5000	0.7111	**0.7955**
	6	0.1698	0.3121	0.6610	0.1560	0.1185	0.1813	0.1552	0.0830	0.2806	0.3153	**0.6720**	0.6477
Recall	1	0.1667	0.2727	0.3939	0.1667	0.2424	0.2424	0.2879	0.0758	0.3485	0.2879	0.4848	**0.5303**
	6	0.0758	0.1247	0.1801	0.0577	0.0508	0.0808	0.0624	0.0439	0.1270	0.0152	0.1940	**0.3756**

Our method gives the best Top-1 Precision out of all the image matching methods

Competitive Top-1 Precision with WhiteNet that has phishing webpages in training pipeline

Conclusion and Future Work

- A visual similarity-based Phishing Detection method that does not require Phishing samples at the training time.
- Surpasses the baseline image-matching methods on detection of phishing with the highest Precision and Recall for Top-1.
- Provides competitive performance to other Triplet Learning based methods (i.e. WhiteNet) that USES phishing samples at the training time.
- Future work includes:

  1. Testing for a larger Phishing webpage dataset
  2. Using lighter and faster shared models for detection
  3. Improvements for structural embedding creation

Thank You